科技期刊语义出版的学术不端防范功能的实现

摘　要：为给科技期刊语义出版的发展及学术不端行为的防范提供参考，根据科技期刊的发展趋势，结合语义出版的特点，研究科技期刊语义出版过程中，防范学术不端行为的实现。在学术不端行为防范方面，语义出版使得科技期刊出版流程主要发生两方面的变化：审稿方面从单一的重复性检测到创新性预测；出版内容方面从单篇论文出版到关联数据出版。由于检测软件的智能化，出版内容的关联化，内容推送的精准化，出版流程的去中心化，科技期刊对于学术不端行为的防范功能得以实现。但是科技期刊的语义出版中还存在数据的开放共享与出版商既有利益的冲突，及语义技术与区块链技术联合应用的冲突，本文就此提出了一些建议。

关键词：语义出版；学术不端；重复性检测；关联出版；区块链
中图分类号：G237　　**文献标识码：**A
文章编号：1671-0134（2019）10-025-03　　**DOI：**10.19483/j.cnki.11-4653/n.2019.10.005

文 / 闻丽　罗列

近年来，学术不端行为呈现了新变化，由于获取信息途径更多，抄袭成本降低，形式越来越多样化，隐蔽性越来越强，逐渐显现常态化。学术不端行为对学术环境的不良影响也越来越受到关注。科技期刊作为科技成果的重要交流平台，是干预和阻止学术不端行为的最后防线，在净化学术环境中占有重要地位。目前，科技期刊出版流程中对于学术不端行为的防范主要是在审稿阶段通过同行评议、检测系统查重、要求作者签署“诚信承诺书”等措施来进行，同行评议和新兴起的查重软件均存在一定的不足，签署科研诚信承诺书基本流于形式，收效甚微，使得学术不端行为向隐蔽性更强、发现难度更大等方向发展[1]。

目前，关于论文出版阶段科技期刊防范学术不端行为的研究主要是集中在学术不端行为规律及其防范措施方面。徐石勇等以《丝绸》为例，对近年来该期刊遇到的作者学术不端的现象进行汇总、分析，并总结了针对这些学术不端现象的防范措施[2]；吴宁对科技期刊中学术不端行为的变化特点进行了分析，并提出了一些应对措施[3]；王子君等提出在修改稿件过程中如何判断学术不端行为[4]。研究表明，由于各途径均存在疏漏，在科技期刊出版中防范学术不端行为最重要的途径还是加强期刊人才建设和提高其素质，不端行为的发现往往需要依赖编辑的责任心和专业素养，需要充分发挥编辑的主观能动性[5-7]，这就难免造成学术不端行为的漏查。

作为一种科技期刊出版的发展业态，关于语义出版的研究多集中于出版流程中语义出版技术的实现，鲜见有关语义出版在防范学术不端方面的研究。本研究基于语义技术在科技期刊的应用，详细、多角度阐述语义出版在技术实施方面如何实现对不端学术行为的防控，旨在为科技期刊语义出版发展中对于学术不端行为的规避提供参考，使得科技期刊成为维护学术生态环境的重要屏障和学术评价的公器。

1. 学术不端行为防范方面科技期刊出版流程的变化

1.1 审稿方面从单一的重复性检测到创新性预测

由于从第一次重复率检测到上网的滞后期，在刊发前编辑可能需对稿件进行多次重复率检测。真正的语义出版工作始于作者进行论文创作，由于文章与检测系统的无缝衔接，确保稿件每次修改都自动生成实时相似性检测，超过预设值则自动发出预警，为编辑工作带来极大的便利。语义出版在稿件的审查阶段，不仅可以防范同行评议、判断创新性中出现学术不端，还可以帮助编辑对于稿件的研究背景、学科进展、应用前景进行预估，不再是单一的进行相似性检测。

应用语义技术有利于基于最少的工作建立一个真正意义上的全网协同平台，实现最大限度的知识产权保护和学术不端行为的发现。从单一的重复性检测到全网协同、实时重复性检测再到创新性判断，语义技术的应用对于尽早发现学术不端行为、对研究内容低水平重复筛选将发挥重要作用。

1.2 出版内容方面从单篇论文出版到关联数据出版

基金项目：中国高校科技期刊研究会青年基金资助（项目编号：CUJS-QN-2018-029）、湖南省教育厅一般项目“语义技术在高校科技期刊发展中的应用”（14C1192）

由于传统出版中的期刊版面限制，作者无法提供数据的分析过程，为数据篡改和伪造提供了便利。为了提高研究的可重复性，科研人员要求期刊提供研究论文原始数据的期望越来越强烈。论文投稿时提交和发表时公布支撑数据，已成为越来越多科技期刊的基本要求，例如 Science、Nature、Cell 等，然而传统出版的有限版面难以满足这一要求 [8]。

2014 年，由 Nature 出版集团创立的 Scientific Data 电子期刊和中国科学院地理科学与资源研究所、中国地理学会创立的“全球变化科学研究数据出版系统”均已实施了实体数据与数据论文关联出版的模式。目前，国内已有《地理学报》等 30 多家期刊就关联数据发表论文达成了共识，已有部分期刊提出了提交支撑数据的要求。如《中华健康管理学杂志》从 2016 年起，逐步要求原始研究类稿件提供相应的原始材料，例如原始数据、原始结果、量表、干预方法、问卷等，但限于传统出版和当前数据出版的局限性，这些资料不能完全发挥其本身的所有潜能。

2. 科技期刊语义出版中学术不端行为防范功能的实现途径

2. 1 检测软件的智能化

通过查重软件检测可以发现部分学术不端行为，但软件查重存在的漏洞和不足也使学术不端行为更加隐蔽，检测出该行为的难度增大。语义技术的应用不仅使得出版内容发生改变，也可使得查重技术更加智能化，查重结果更为准确。目前，语义技术是查重软件应用的技术之一即是基于词频统计的方法。此方法引自向量空间的检索模型，和信息检索技术相关。首先要统计每篇文档中各个单词的出现次数，再根据指定规则将单词频度转化为空间特征向量，最后采取度量向量之间的距离来计算相似度，从而达到查重的目的 [9]。

近年来，中国知网开发的科研诚信管理系统实现了基于内容的信息指纹技术与语义分析技术的有效结合，不仅保证了检查结果的精准性，还具备了较强的抗干扰能力，支持文献改写、重组、翻译等多种变换形式的检测，检测系统已实现了语义级别内容的检测。该系统不仅可对图、表等特殊检测对象进行基于标题、上下文、图表内容结合的相似性检测处理 [10]，还可根据特定的概念、观点、结论等内容进行智能信息分类处理。随着语义技术在查重系统中应用的逐步成熟，粒度可调地融入语义特征的相似性分析技术，以句子为基本分析单位，专门用于文本相似性的快速准确比对。其强大的查重功能将会加大学术不端检测力度，即使是一般的低水平重复也会被检测出来，普通的逃避查重行为更是无处遁形，从而在剽窃和抄袭方面的学术不端行为防范中发挥重要作用。

另外，仅根据相似比来判断是否存在学术不端行为，决定稿件“去”与“留”有着一定的不科学性。要真正地减少跟风、重复选题，遏制平庸不良选题，不仅要杜绝“形抄”，更要杜绝“意抄”。应用语义技术的科技期刊出版，将呈现智能化出版模式。智能化的信息获取方式，可以让电脑根据已有信息和用户的偏好，直接获得完成分析结果。

2. 2 出版内容的关联化

通过碎片化论文内容加上关联原始数据信息，可以完整再现作者的研究过程，数据来源更清晰，文章质量更可信，并可实现对论文内容的再分析，客观评价其对科学发展的价值和影响。复证是检验学术出版内容真实性和科学性的有效手段，通过使用论文中数据和方法进行重复操作，将试验结果与由获得原始数据进行比较，判断学术成果的真实性、可信性和可靠性。利用语义技术进行关联数据出版，可完整展示试验过程，提供复证基础，增大了不端行为被发现的可能性。通过语义出版共享这些资源，还可让更多的数据资源使用者从多个方面主动参与论文“真伪”的辨别。虽然不能完全排除数据篡改和伪造的可能性，但是造假本身所需付出的成本被扩大。

2. 3 内容推送的精准化

大量的网络信息被不加选择地广泛推送给不同的用户。一般的主动性推送，在引起用户焦虑的同时也容易被忽略。而量体裁衣的精准推送可以解决这一问题。通过分析用户的偏好库，语义出版可以智能地将出版内容主动推送给相关领域的专家和学者，这样“被作者”和“被免冠作者”的情况就会被及时发现。由于专业所限，一般的科研圈子范围有其有限性，语义出版的精准智能推送增大了抄袭行为对于目标对象的曝光概率，将使得此类情况无处遁形。

2. 4 出版流程的去中心化

现行的版权保护模式已经不能适应时代的需求，一纸简单的“诚信承诺书”，也仅是道德层面的约束。语义出版实践的数据发布过程中采取的是 FAIR 原则，这一原则与去中心化所遵循的原则有较大的重叠，即数据存储独立于服务提供商、具有交互性、可访问性以及自由表达和处理网络信任的能力 [11]。语义出版天然具有去中心化的特点，而去中心化可以防止文章被任意篡改。真正的语义出版是在论文创作阶段就开始了语义分析和处理 [11]，如果语义出版自研究人员从试验数据的记录开始，其后续发生数据篡改、版权争议等事件的概率将大大降低。

3. 科技期刊语义出版中学术不端行为防范功能的实现障碍

3.1 数据的开放共享与出版商既有利益的冲突

单一的重复性检测不足以充分发挥语义技术的优势。学术不端检测需要一个尽可能完备的全文数据比对资源库。目前，中国知网的科研诚信管理系统不仅可以与知网所提供的《中国知识资源总库》内所收录文献的比对，还可与互联网部分开放的资源进行比对，但是对于其它数据库的资源无法涉及。语义出版与数据的开放共享是互利的，可以实现学术不端检测的全网协作，但是这与各数据库出版商的利益存在冲突。

3.2 语义技术与区块链技术的联合应用的冲突

语义出版是有限的去中心化，通过与其他技术联合使用，设计具有可信任性的语义出版物，实现完全去中心化，如区块链技术。区块链是通过密码学方式形成的一个由集体维护的分布式数据库，数据块相连组合成链条，各节点互相监督，整个工作流程中移除了传统第三方，信息互联互通而又无法被篡改，安全性高[12]。通过区块链技术可以对内部数据块进行版权登记，并且其共识机制使版权交易公开透明[13]。虽然区块链技术在版权方面的运用还在探讨与尝试阶段，但其不失为知识产权保护的有效方法之一。王众等就两项技术在农产品质量安全追溯领域的联合应用提出了一个概念性的验证系统[14]。在语义出版的基础上应用区块链技术，通过智能合约的方式建立共识机制，保障数据获取的安全性和权益归属，为科技论文出版添加上信任的纽带和数据安全的保障。语义技术是使得信息可用，而区块链技术是确保隐私不被非法获取和数据安全，二者间存在着一定的冲突。虽然区块链技术与语义技术的融合对于防范学术不端颇具效力，但其可行性还需验证。

4. 结论和建议

语义出版是具有弹性扩展、动态分配和资源共享等特点的数字内容出版，它催生大量新型出版业信息服务，带动出版格局的整体变革。充分利用语义技术强大的语义分析功能，可提高查重检测的精确度，通过语义出版从多方面提高论文学术不端行为的辨识率，从技术手段方面防范学术不端，有利于对科技期刊的出版流程进行更为科学的控制和管理，建立科技期刊知识版权的保护壁垒，从而确保科技期刊的绿色健康发展，建立良好的学术环境。

在开放科学的大背景下，出版数据的开放共享趋势明显。在语义出版中，可以实施出版数据的开放共享与部分有偿使用。在进行学术不端检测时，通过与各出版商的利益关联，综合性无偿服务与个性化有偿服务相结合，提高其共享数据的积极性，从而实现全网协同检测。

目前，内容组织的语义化已成为大数据时代科技期刊发展的主要趋势。但是要使广大的作者及编辑人员熟练地应用语义技术来写作和编辑论文，还须开发出友好的软件操作界面以供使用。各刊社要重视专业人才的培养及现有人员知识的更新，有条件的单位可考虑引进相关专业技术人才，已形成集群的科技期刊则可采取远程的专业智力资源共享。媒

参考文献

[1] 王文福 . 网络时代期刊论文隐形学术不端挖掘策略 [J]. 中国科技期刊研究，2016，27（7）：677-682.

[2] 徐石勇，叶靖，康锋，等 . 期刊学术不端的现象、成因及防范措施 [J]. 编辑学报，2019，31（4）：411-414.

[3] 吴宁 . 科技期刊中学术不端行为的演变及应对措施 [J]. 编辑学报，2019，31（4）：369-371，376.

[4] 王子君，赵丽琴 . 修稿过程中甄别学术不端的问题研究 [J]. 科技传播，2019，11（13）：146-147.

[5] 陆宜新 . 学术不端检测中的隐性重复现象分析 [J]. 中国科技期刊研究，2015，26（6）：552-555.

[6] 彭芳，金建华，董燕萍 . 同行评议造假原因分析及防范措施 [J]. 编辑学报，2018，30（3）：240-243.

[7] 徐海丽 . 学术不端行为及其预防措施 [J]. 中国科技期刊研究，2015，26（6）：545-551.

[8] 孔丽华，刁妍，郎杨琴，等 . 数据期刊中科学数据的同行评议方法研究 [J]. 编辑学报，2019，31（3）：262-266.

[9] 邢娜 . 查重技术及其在信息校验中的应用研究 [D]. 天津：南开大学，2008.

[10] 孙雄勇 . 学术不端文献检测系统在学术期刊中的应用情况分析报告 [N]. 中国新闻出版广电报，2016-12-30（8）.

[11] 徐雷 . 几个值得关注的语义出版研究热点 [J]. 出版科学，2018，26（2）：16-22.

[12] 孙婧，管青山，段立晖，等 . 科技期刊媒体融合出版现状与数字版权保护及建设思考 [J]. 中国科技期刊研究，2018，29（8）：813-821.

[13] 牛敏 . 基于区块链技术的数字版权管理模式研究 [D]. 北京印刷学院，2017.

[14] 王众，梁早清，郑业鲁，等 . 语义网技术与区块链技术的交集在农产品质量安全追溯领域的应用 [J]. 农业图书情报，2019，31（1）：60-68.

（作者单位：中南林业科技大学期刊社）